

Optimization modeling for analyzing fantasy sport games

J. Beliën*, D. Goossens** and D. Van Reeth***

*Center for Information management, Modeling and Simulation, Faculty of Economics and Business, KU Leuven campus Brussel, Warmoesberg 26, 1000 Brussels, Belgium, jeroen.belien@kuleuven.be,

** Faculty of Economics and Business Administration, Ghent University, dries.goossens@ugent.be

*** Research Centre for Globalization, Innovation and Competition, Faculty of Economics and Business, KU Leuven campus Brussel, Warmoesberg 26, 1000 Brussels, Belgium, daam.vanreeth@kuleuven.be,

Abstract. In a fantasy sport game, participants act like a team manager building a team of real individual players of a professional sport. The real performances of these players (or their teams) are translated into points for their team managers. The managers' aim is to collect as many points as possible thereby defeating the fantasy teams of opponents. First, we discuss a number of common game rule characteristics of fantasy sport games. Based on these characteristics, we present a mixed integer programming model to produce and analyze ex-post results for a large variety of fantasy sport games. We discuss how these results create value for both the game organizer and the participants. Finally, we apply our system in practice to a fantasy cycling game.

Keywords. Fantasy game, optimization, ex-post, game rules

1. Introduction

A fantasy sport game allows its participants to act as a team manager building a team of real individual professional sport athletes. The real-world performances of these athletes (or their teams) are translated into points for their team managers. The managers' aim is to collect as many points as possible thereby defeating the fantasy teams of opponents. In the remainder of this paper we will use the word 'participant' to describe a virtual team manager while the word 'player' solely refers to a real-world athlete.

Fantasy sports found their origin in the United States in the 1980s. American journalists Glen Waggoner & Daniel Okrent developed a game in which a handful of participants would draft from a pool of active baseball players (Davis & Duncan, 2006). Fantasy sports really took off around the turn of the millennium, when the Internet transformed fantasy sports into a mainstream phenomenon. Fantasy sports are now being played by tens of millions of people worldwide. Fantasy American Football has become the most popular fantasy sport league, with a market share of 80% in the United States and Canada (see www.fantasysportsadnetwork.com). In Europe, soccer is the most popular fantasy sport subject. The enormous success of fantasy sports has a lot to do with the fact that it allows online participants to assume an active role of a team owner or a team manager in a sport they are heavily interested in, thereby intensifying the way live sport is consumed. As a result, many sports enthusiasts are now obtaining their sports entertainment through fantasy sports rather than solely by watching games on television (Nesbit & King, 2011).

This paper presents a mixed integer programming model for finding an optimal set of decisions in fantasy sport games under the assumption that all data are known beforehand. Thus, the system allows to identify ex post an optimal team selection and player transfers. One could ask why an ex-post analysis for fantasy sports is useful. After all, participants need to select their team and make transfers before this information becomes available. To describe the value of this ex-post analysis we distinguish between the *information value* and the *commercial value* of the results.

First, our system can be used to create information value. Fantasy sports are often organized by media groups whose core business is to provide information and sell news. On a regular basis they publish the fantasy sport game results and rankings through a website or newspaper. The related articles are often limited to an interview with the week winner and an overview of the top-x best performing teams of the specific week, and the updated general classification. Many readers, however, would be interested in more detailed

statistics and analyses, providing answers to questions like “what is the optimal team of this week?”, “what is the optimal team so far (with respect to different objectives such as maximal price money, maximal number of week victories, etc.)?”, “what were the optimal team changes for the most recent transfer period?”, etc. These kinds of questions can easily be answered by the ex-post optimization model. Moreover, the system can also be used to provide individual feedback to participants (e.g., through a personal website account) that specifically concerns the participant’s team. For instance, given the participant’s starting team, which were the optimal transfers with respect to a particular game objective? What was the highest ranking that could be achieved (at this moment)? Our model can thus be used to generate information that goes much further than only providing the participant’s current ranking. We are convinced that the availability of this detailed information can be a distinguishing factor for choosing between competing fantasy sports.

Second, the results of applying our system also have commercial value. It is our experience that an optimal strategy largely outperforms that of the best participant. Having this in mind, last season’s optimal team can be used for the promotion campaign of the new season, highlighting the (hopefully large) difference with the score of the best participant, who won the competition. By comparing these scores, people tend to overestimate their chances of winning the game. Consequently, the game can attract more participants which increases the turnover for the organizer. Another example of how our system creates business value is by using the results to stimulate people to join the game even after it has already been started. Indeed, most fantasy sport games allow people to submit a (new) team at any time during the game (as long as they pay the participation fee). However, these late participants have to start with a handicap as the other participants have already had opportunities to collect points. That is where our system comes into play. By highlighting the difference between the last season’s optimal score (calculated by our model) and the score of the last season’s winner, one can possibly convince doubting people of still having a fair chance of winning the general classification. This will certainly be the case if this difference is (substantially) higher than the total score of the current leader. After all, new participants also have an important advantage as they face less uncertainty concerning, for instance, the players’ expected performance. Moreover, the ex-post optimal team often looks so familiar that people get the (false) impression that this optimal team could easily be predicted beforehand or that it was, at least, not that difficult to compose a team that would have beaten the current round (e.g., race or matchday) winner. Consequently, by regularly publishing the optimal team for every round and the difference with the round winner, people will be continuously stimulated to submit a new team that aims at winning a specific round.

The remainder of this paper is organized as follows. Section 2 provides a literature review and a number of common characteristics of fantasy sports. Based on these common characteristics, Section 3 presents a mixed integer programming (MIP) model that can be used to analyze the ex-post results of fantasy sport games. We discuss how we applied our model on a real-life fantasy cycling game in Section 4. Finally, Section 5 concludes this paper and lists directions for future research.

2. Fantasy sport games: an overview

2.1 Literature review

The growth of fantasy sports has made it an important part of the sports industry. The booming popularity also stimulated research on fantasy sports and from 2005 on, literature on fantasy sports really started to grow. We discern two lines of research.

The first line is economically oriented. It sees fantasy sports as a new form of sport consumption and studies how this affects the behavior of sports fans. Randle & Nyland (2008), Drayer et al. (2010) and Karg & McDonald (2011) all take a rather global view and look at the impact of fantasy sport participation on the various media sources fantasy sport participants use. The specific impact of fantasy sport on television ratings is analyzed by Nesbit & King (2010), Dwyer (2011a) and Fortunato (2011), while the live attendance impact is taken up by Nesbit & King (2011). Most of these studies basically conclude that instead of competing with traditional ways of sport consumption, fantasy sport appears to be a complementary and value-adding activity (Dwyer, 2011a). Fan loyalty and how fantasy sports can be used in customer relationship management are studied by Dwyer (2011b) and Smith, Synowka & Smith (2010).

A second research angle analyses sociologically oriented aspects of fantasy sport participation. Participants in fantasy sports are predominantly young, white, male and well-educated. For instance, based on a web-based survey of nearly 1,200 respondents Levy (2009) states that 98% of the fantasy sports enthusiasts in the US are male, 94% are white, and 69% have at least a bachelor's degree. The Fantasy Sports Ad Network (www.fantasysportsadnetwork.com) comes to a similar conclusion. They claim that 92% of fantasy participants are male, 91% are white and 71% have a bachelor's degree or higher. According to them, the average age of a fantasy sport participant is 36. Some of the studies focus on the reasons why sports fans participate in fantasy sports. Often cited and validated motivational dimensions are fanship, entertainment/escape, competition and social interaction (see, for instance, Davis & Duncan, 2006; Farquhar & Meeds, 2007; Dwyer & Kim, 2011; Ruihley & Hardin, 2011; Lee, Seo & Green, 2012; Billings & Ruihley, 2013). Although the conclusions on the importance of some of these dimensions differ from one study to another, the results generally suggest a pattern of fantasy sport participation that is more purposeful and active than traditional media use (Dwyer & Kim (2011), p. 70).

Perhaps surprisingly, decision support for fantasy sport games has not received much attention in the literature so far. There are a number of papers on mathematical techniques to support coaches with team formation in real sports, such as Sierksma (2006), Tavana et al. (2013), Dadelo et al. (2014). However, to the best of our knowledge, only work by Bonomo et al. (2014) addresses decision support for fantasy sport games. These authors study Gran DT, a fantasy game on the first division of the Argentinian soccer league. In order to support the participants, they present an a priori model, which creates a competitive team, as well as an a posteriori model, which determines what would have been an optimal team composition once all the results are known. Our work aims to contribute to this line of research, by developing an optimization model that can be applied to many (if not all) fantasy games, and by providing decision support not just for the participants, but also for the organizers.

2.2 Characteristics

To our knowledge, there are no overview studies that list fantasy sports or classify them according to a certain set of characteristics. Such a work would, in fact, be virtually impossible since the number of fantasy sports is extremely large, even if one would focus on a single sport. Based on observation, we list a number of game rules characteristics that are fairly common in fantasy sports (see Table 1). The fantasy sport games listed in the rows of Table 1 are not exhaustive, but merely a selection of representative fantasy sport games for different sports disciplines. The game rules characteristics listed in the columns of Table 1 give, to the best of our knowledge, a complete image of all possible game rules.

First, team selection in fantasy sport can either be subject to a budget constraint or can be free from such a constraint. Each player that can be selected has a given value. This value is based on past performances of the player and can either be a fictitious monetary value (e.g., based on his salary or transfer price) or a sports score (e.g., ATP points in tennis). Most games usually have a budget constraint to make it impossible to select only the best (usually high-valued) players. The remaining budget, that is the budget not spent on players, is often used as a tiebreaker in case several participants end up with the same score.

Second, some fantasy sports allow participants to change their fantasy team selection at well-defined moments during the season while in other fantasy sports the team selection is fixed from the start for the whole of the season. Of course, when it is possible to transfer players during the season, the game becomes much more strategic thus creating an even more intense need and desire to follow the sport closely, which is to the benefit of the media company that is behind the game. In Table 1, the number of transfer opportunities is listed in format axb , where a represents the number of transfer periods and b the maximal number of transfers in each period. In some games, the number of transfers are unlimited but each transfer 'costs' a certain number of points. When transfers can be made during the game and the decisions made in earlier periods might affect the decisions made in later periods (e.g., through the remaining budget), we call the fantasy sport game *dynamic*. Otherwise, it is a *static* fantasy sport game.

Third, while in many fantasy sports all players of a fantasy team earn points for the participant, there are also fantasy sports for which this is not the case. For instance, some soccer fantasy sports allow participants to select a basic team of 11 players and a number of substitutes. Only in case a basic player does not play on

a certain matchday, the substitute can earn points for the participant. Another possibility is that only the x highest-scoring players in the fantasy team earn points for the participant on each matchday or race.

Table 1: Game rules characteristics of fantasy sports

Game	Limited budget	Value for remaining budget	Number of transfer opportunities (transfer periods x number of transfers)	Points contributed by players may differ from team to team	Number of players to select from	Team size	Upper bound on number of different teams	Unrestricted team formation	Players can be selected in different teams (no draft)	Number of rounds	Intermediate prizes
Gran DT (Bonomo et al., 2014)	yes	no	14x4	yes	500	18	4.37E+32	no	yes	15	yes
Fantacalcio (www.fantacalcio.kataweb.it)	yes	no	37x18	yes	559	18	3.36E+33	no	yes	38	yes
Fantasy Baseball MLB (www.mlb.com/mlb/fantasy/fb/info/index.jsp)	no	no	16x24	yes	1,200	16	7.99E+35	no	no	24	no
English soccer's fantasy premier league (fantasy.premierleague.com)	yes	no	unlimited ¹	yes	400	15	6.29E+26	no	yes	38	yes
Dunkest, the online American Fantasy Basketball game (www.dunkest.com)	yes	no	unlimited ¹	yes	600	11	8.29E+22	no	no	44	no
Fantasy cycling (fantasy.road.cc) Tour de France	yes	no	45	yes	200	9	1.18E+15	yes	yes	22	yes
Fantasy Velogames (www.velogames.com) Tour de France	yes	no	0	no	200	9	1.18E+15	no	yes	22	yes
Gigabike (http://www.gigabike.be)	yes	yes	5x5	yes	3,150	30	2.92E+72	yes	yes	19 0	yes

¹Transfers at the cost of points

Fourth, each fantasy sport game has a database of players out of which participants have to select a team of a certain team size. Calculating the number of possible combinations of the required team size out of the total number of players selectable gives a theoretical upper bound on the number of different teams possible. This theoretical number is may be a considerable overestimate if team selections have to satisfy some rules, like the already discussed budget constraint. Next to the budget constraint, many fantasy games have extra rules with respect to team formation. For instance, in a football fantasy game, a team might need to have 1 goalkeeper, 4 defenders, 4 midfielders and two strikers. Moreover, in some fantasy sports players can only be selected once, called a draft. A draft usually goes in turns with each participant picking one player per turn. If one participant picks a given player for his fantasy team, the player becomes unavailable to the other participants. Of course, this reduces significantly the number of people that can participate in the game. Therefore, in most fantasy sports, players can be selected infinitely, this means by every participant independently of the decisions of the other participants. In this way, all participants have equal chances of winning and the game organizer can allow a much larger number of participants.

Finally, a fantasy sport game usually consists of several playing rounds (e.g., matchdays in football, races in cycling). Every round consists of one or more events (games) for which points can be collected by the fantasy team. Some fantasy sports only have prizes at the end of the competition while others also have day

prize winners for each round or for sets of rounds (e.g., each month). With regular prizes throughout the season, sport loyalty and game affiliation are further increased, again very much to the benefit of the media that are behind the game. Moreover, this offers an extra incentive for participants to enter the fantasy game even after the start of the competition.

3. Optimization model

This section presents a mixed integer programming model for finding optimal ex-post decisions in fantasy sport games (optimal start selection as well as optimal transfers). In other words, the model allows to identify an optimal strategy if one knows beforehand how many points each player will collect in each game.

We define the following sets and indices:

$p, p' \in P = \{1, 2, \dots, P \}$	players (athletes, riders, etc.),
$\pi \in \Pi = \{1, 2, \dots, \Pi \}$	player types (e.g., defenders, attackers, midfielders in football; leaders, domestiques in cycling, etc.),
$P_\pi \subseteq P$	set of players of type π
$t \in T = \{1, 2, \dots, T \}$	periods; between each pair of periods the fantasy team can be modified,
$\tau \in \Theta = \{1, 2, \dots, \Theta \}$	games, i.e., events where players can gain points (matches, races, etc.),
$G \subseteq \Theta$	sets of games that are considered as a whole for assigning prices, e.g., matchdays, months, cycling tours, etc.,
$t_G \in T$	period in game set G that is used as a reference for the remaining budget,
$e \in E_G = \{1, 2, \dots, E_G \}$	ranks for which a prize is obtained in game set G (e.g., first, second, third, etc.).

The model parameters are as follows:

$v_{p\tau}$	points obtained by player p in game τ ,
L	points lost per used transfer opportunity
c_{pt}	cost of player p in period t ,
B	available budget,
$n_{\pi t}$	minimal number of players of type π that have to be selected in period t ,
$N_{\pi t}$	maximal number of players of type π that can be selected in period t ,
A_t	maximal number of transfers allowed in period t ,
D_τ	maximal number of players that can earn points for the team in game τ ,
$a_{p\tau}$	= 1 if player p played in game τ ; 0 otherwise,
ω_{eG}	value (e.g., prize money) of obtaining rank e in game set G ,
H_{eG}	score of the opponent that obtained rank e for game set G ,
R_{eG}	remaining budget in period t_G of the participant that obtained rank e for game set G ,
Q_π	maximal number of players of player type π in the player pool,
ε	weight of the tie-breaker criterion.

The decision variables are as follows:

x_{pt}	= 1 if player p is in the team in period t ; 0 otherwise,
$y_{p\tau}$	= 1 if player p earns points for the team in game τ ; 0 otherwise,
r_t	remaining budget in period t ,
z_{pt}	= 1 if player p is transferred into the team in period t , 0 otherwise,
s_{pt}	= 1 if player p is a substitute in period t , 0 otherwise,
w_{eG}	= 1 if the team obtains rank e for game set G , 0 otherwise,
q_p	= 1 if player p is included in the player pool.

The optimization model can then be formulated as follows.

$$\text{Maximize } \sum_{p \in P} \sum_{\tau \in \Theta} v_{p\tau} y_{p\tau} - L \sum_{p \in P} \sum_{t \in T} z_{pt} + \varepsilon r_{|T|} \quad (1)$$

$$\text{s.t. } \sum_{p \in P_\pi} x_{pt} \geq n_{\pi t} \quad \forall \pi, \forall t \quad (2a)$$

$$\sum_{p \in P_\pi} x_{pt} \leq N_{\pi t} \quad \forall \pi, \forall t \quad (2b)$$

$$\sum_{p \in P} c_{p1} x_{p1} + r_1 = B \quad (3)$$

$$\sum_{p \in P} \sum_{t \in T} c_{pt} (x_{pt} - x_{p,t-1}) = r_{t-1} - r_t \quad \forall t > 1 \quad (4a)$$

$$z_{pt} \geq x_{pt} - x_{p,t-1} \quad \forall t > 1 \quad (4b)$$

$$\sum_{p \in P} z_{pt} \leq A_t \quad \forall t > 1 \quad (4c)$$

$$y_{p\tau} \leq x_{pt} \quad \forall p, \forall t, \forall \tau \in T_t \quad (5)$$

$$\sum_{p \in P} y_{p\tau} \leq D_\tau \quad \forall \tau \quad (6)$$

$$\sum_{p \in P_\pi} s_{pt} = 1 \quad \forall \pi, \forall t \quad (7a)$$

$$s_{pt} \leq x_{pt} \quad \forall p, \forall t \quad (7b)$$

$$y_{p\tau} \leq (1 - s_{pt}) + \sum_{p' \in P_\pi \setminus p} (1 - a_{p'\tau}) x_{p't} \quad \forall \pi, \forall p \in P_\pi, \forall t, \forall \tau \in T_t \quad (7c)$$

$$x_{pt} \in \{0,1\} \quad \forall p, \forall t \quad (8)$$

$$y_{pt} \in \{0,1\} \quad \forall p, \forall t \quad (9)$$

$$z_{pt} \in \{0,1\} \quad \forall p, \forall t > 1 \quad (10)$$

$$s_{pt} \in \{0,1\} \quad \forall p, \forall t \quad (11)$$

$$r_t \geq 0 \quad \forall t \quad (12)$$

The objective function (1) maximizes the points collected over all games, while taking into account a penalty for each used transfer opportunity. In case several solutions exist with maximal points, a tie-breaker criterion should be applied. We opted for selecting the solution with the highest remaining budget in the final period, which explains the third term in the objective function. The model can easily accommodate other tie-breaker criteria, such as time of registration, highest rank obtained in some particular round, number of goals scored by selected players, etc. Constraint sets (2a) and (2b) ensure that in each period the number of players of each type is within the required bounds. Constraint set (3) models the budget restriction for the first period. If it concerns a static fantasy game, constraint (3) suffices.

For a dynamic fantasy game, the remaining budget of a preceding period can often be used as additional budget in the succeeding period. This is modeled by the constraint sets (4a). The cost of the new players in the team, i.e., players for which $(x_{pt} - x_{p,t-1})$ equals 1, must be smaller than the cost players transferred out, i.e., players for which $(x_{pt} - x_{p,t-1})$ equals -1, plus the remaining budget of the preceding period, r_{t-1} . The difference is the remaining budget r_t of period t . In this scenario, the maximal number of transfers between two periods may be limited. This is handled by the constraint sets (4b) and (4c). Constraint set (4b) ensures that the binary variable z_{pt} is forced to 1 if a player is transferred into the team. Constraint set (4c) guarantees that the maximal number of transfers is not exceeded.

Constraint set (5) models the scenario in which not all selected players automatically earn points for the fantasy team. First, in some fantasy sports, only the points of the D_τ best performing players in each game τ

are counted. This is modeled by constraint set (6). Second, in order to avoid that participants lose their interest in case of inactivity (e.g., caused by injury) of a selected player, some fantasy sports require participants to select a substitute player for each player type. This substitute only earns points for the team if at least one of the other players of the substitute's type did not come into action. For instance, in some fantasy football games, participants select four basis midfielders and one substitute midfielder. The substitute does not earn points in a particular game unless (at least) one basis player did not play that game, even in the (unlucky) case that the substitute scored more points than the basis player. The constraint sets (7a)-(7c) model the issue of substitute players in fantasy sports. Constraint set (7a) ensures the selection of one substitute player for each player type. Constraint set (7b) guarantees that the substitute player is one of the selected players. Constraint set (7c) ensures that the substitute player can only earn points if at least one of the basis players did not play. Finally, the constraint sets (8)-(12) define the variable domains.

Model (1)-(12) can be used to find the team that would have scored the maximal number of points for every fantasy sport game that we are aware of. Objective function (1) can be easily adapted for evaluating alternative goals. For instance, if one is interested in the team that scored the most points in particular classes of games (e.g., mountain stages in cycling), only the points scored in those games are retained in the objective function. Some fantasy game variants, however, require an alternative model. They are listed hereafter.

In a first alternative scenario, the participants make a start selection of players, referred to as the player pool, out of which they select subsets of players for particular (sets of) games. This should better reflect the decisions made by a real team manager, who has to compose an extensive team at the start of the season and makes a team selection for every game. To model this scenario the constraint sets (4a)-(4c) are replaced by the constraint sets (13)-(15). Constraint set (13) ensures that in future periods only players of the player pool can be selected, while constraint set (14) restricts the number of players of each type in the player pool. Constraint set (15) defines the domain of q_p .

$$x_{pt} \leq q_p \quad \forall p \forall t \quad (13)$$

$$\sum_{p \in P_\pi} q_p \leq Q_\pi \quad \forall \pi \quad (14)$$

$$q_p \in \{0,1\} \quad \forall p \quad (15)$$

Second, it often occurs that one wants to know the team that would have won the most prizes (e.g., to find the team that maximizes the prize money). To this aim, the objective function is rewritten as (16) and the constraint sets (17)-(19) are added.

$$\text{Maximize} \quad \sum_G \sum_{e \in E_G} \omega_{eG} w_{eG} \quad (16)$$

$$\text{s.t.} \quad \sum_{p \in P} \sum_{\tau \in G} v_{p\tau} y_{p\tau} + \varepsilon r_{t_G} - w_{eG} (H_{eG} + \varepsilon R_{eG}) \geq 0 \quad \forall G, \forall e \in E_G \quad (17)$$

$$\sum_{e \in E_G} w_{eG} \leq 1 \quad \forall G \quad (18)$$

$$w_{eG} \in \{0,1\} \quad \forall e, \forall G \quad (19)$$

Objective function (16) maximizes the number of weighted game set winning ranks obtained by the team. A game set is defined as a set of games for which a prize can be won. The weight ω_{eG} is typically the prize money associated with rank e in game set G . For instance, in many fantasy games, prizes are assigned on a periodical basis. Constraint set (17) ensures that rank e is obtained for a particular game set G ($w_{eG} = 1$) only if either the team's score is higher than the score of the opponent that obtains rank e or the team obtains a score equal to the opponent's score, but has a larger tie breaker value (i.e. remaining budget) in the corresponding period. Notice that for ease of notation, we have dropped the penalty for each used transfer opportunity here. Constraint set (18) guarantees that a team can obtain at most one winning rank for each

game set. This model assumes that higher ranks entail a higher value ω_{eG} which seems to be a realistic assumption. Finally, constraint set (19) defines the domain of w_{eG} .

From a computational point of view, this second alternative scenario can be challenging. The formulation can however be strengthened with valid inequalities of the following format for any rank e in a game set G and a set of player-period pairs Q .

$$w_{eG} \leq \sum_{(p,t) \in Q} y_{pt} \quad (20)$$

Each set of player-period pairs Q needs to be chosen such that one needs at least one player p from $(p, t) \in Q$ in the team in period t in order to have a possibility of obtaining rank e for game set G . These sets can be generated using a binary search tree, where each level corresponds with the choice whether or not to include some player in the team for some period. Player-period pairs that contribute more to the scored points are handled earlier in the tree. In each node, we can compute an upper bound on the points that can be scored, given the decisions made in the search tree so far. In order to limit the computation time, we can ignore constraints with respect to the composition of the team or the budget when computing upper bounds (although of course, this will weaken the bound). As soon as the upper bound in a node drops below the score of the best participant, we can trace back the pairs (p, t) we excluded on the path to the root node in order to construct a set Q . Indeed, we know that at least one of these player-period pairs must be part of the team line-up in order to beat the score of the participant that obtained rank e for game set G . Notice that there is no point in further exploring the branches of the tree starting from this node, as subsequent sets Q would only create valid inequalities which are dominated by the ones we already found.

As most of the fantasy sports that we are aware of, if not all, entail a simpler game concept than the rules incorporated in model (1)-(12), (13)-(15) and (16)-(19), our model can be used for a large variety of fantasy sports. For instance, the ex-post optimization model for a static fantasy sport, in which teams are fixed during the whole length of the game, is less complicated than the model for a dynamic fantasy sport. As a matter of fact, the static model can be seen as a special case of a dynamic model with only one period. Consequently, our model can be used for the ex-post optimization of both static and dynamic fantasy sports. A similar reasoning applies to the characteristic of multiple (daily) prizes. Fantasy sports with only one end prize can be seen as a special case of the situation with multiple prizes. Consequently, model (16)-(19), which involves multiple winning ranks, can be used for both types of fantasy sports. The same holds for the characteristic that only a limited number of players per team yield points. Fantasy sports in which all selected players earn points on every matchday or race can be seen as a special case in which the number of players that yield points equals the team size. Finally, also the use of the remaining budget as a tiebreaker leads to a model extension as compared to the case in which the remaining budget is neglected. The latter can be seen as a special case in which the weight of the remaining budget (ε) is set to 0 in the objective function.

4. Practical application: Gigabike

Gigabike is a fantasy cycling game that is hugely popular amongst Belgian and Dutch professional road cycling fans. The Gigabike population corresponds quite closely to the typical fantasy sports participants as reported by Levy (2009). In the 2014 edition of the game, according to the organizers, 96,7% of the participants are male, all are white and they have an average age of 38 years. No information on the educational level of the Gigabike participants was available though.

The game basically operates as follows. At the start of the road cycling season, each Gigabike participant selects a team of 30 riders from a world ranking of all professional road cyclists. In this ranking, each cyclist is assigned a so-called *cycling quotient* (CQ) value based on his performance during the last 12 months (see <http://www.cqranking.com>). This value reflects the past quality of the rider and might indicate his future performance. The total CQ value of a team must be less than a given CQ budget of 20,000. At each of five fixed moments during the season, one can modify the team by substituting at most five cyclists. Hence, the season consists of six periods. However, the sum of the CQ values of the incoming riders cannot exceed the

sum of the values of the outgoing riders plus the remaining CQ budget of the preceding period, if any. There is some incentive to have some remaining budget, as it is used as a tie breaker. For each race, only the eight best scoring riders of the team yield points. The winner of the game is the person whose team gains the most points over the course of the whole season. Table 1 also makes clear how the fantasy sports characteristics apply to the Gigabike game.

We illustrate our approach on the Gigabike game. Comparing the Gigabike's game rule characteristics with other fantasy sports' game rule characteristics (see Table 1), we conclude that Gigabike is the most complex and complete fantasy sport game that we are aware of. It makes use of a constrained budget, in which the remaining budget is used as a tiebreaker, there are transfer opportunities, the remaining budget is transferred to future periods (dynamic game), and the points contributed by players may differ from team to team. Moreover, of all fantasy games that we are aware of, Gigabike has the largest team number (30) to be selected from the largest database of players (3,150 riders). Therefore, the theoretical upper bound on the number of different teams is $2.92E+72$. Of course, due to the budget restriction, the real number of different teams is much smaller, but, compared to other fantasy sport games, Gigabike has no limits with respect to team formation. In Gigabike, there are no player types (e.g., attackers, midfielders, defenders in football) and a single cyclist can be in many different fantasy teams. Consequently, all combinations of riders can be selected as long as the budget constraint is satisfied. In conclusion, of all fantasy sport games we are aware of, Gigabike not only has by far the highest upper bound on different teams, but it also has the largest number of events (190 race days) in which points can be earned. Hence, from a computational point of view, Gigabike poses a demanding test, compared to most other fantasy sport games. Finally, Gigabike offers a broad range of different daily prizes and overall prizes.

Gigabike does not penalize the number of transfer opportunities taken ($L=0$), so the objective function (1) is simplified slightly. Furthermore, Gigabike does not involve substitute players which means that constraints (7a-7c) and (11) do not apply. Finally, participants need not select a player pool to choose from during the season. Hence, the first alternative scenario (13-15) is not relevant in this case. All other constraints are necessary.

4.1 Computational experiments

At first sight, when we focus on a single period and ignore transfers, Gigabike bears some resemblance to the well-known knapsack problem. Indeed, in each period, a number of riders needs to be selected. Each rider collects points and has a cost, and the total cost of the selected riders must fall within a given budget. If we relax the rule that says that only the eight best scoring riders in each race contribute to the team score, and count all points scored by the selected riders instead, the resulting problem is a knapsack problem¹. Experiences with this highly simplified setting for teaching purposes are described in Beliën et al. (2011).

We used a Windows-based laptop with an Intel Core i7 3GHz processor and 8 GB of RAM, running IBM Ilog Cplex 12.6 for our computational experiments to solve several editions of Gigabike. All computation times fell within 20 minutes, though usually less than 2 minutes were sufficient. Nevertheless, we wanted to get an idea of the impact of the problem size on the computation time. Therefore, we varied the number of riders and the number of transfer periods, as these are the parameters that are included in nearly every fantasy sport game. Gigabike has about 1,000 riders each edition that score points. Since it is not obvious to realistically simulate extra riders, and 1,000 scoring players is already a lot compared to most other fantasy sport games, we chose to look at settings with less riders: 750, 500, 250, and 100 riders were randomly picked from the scoring riders. The average computation times over 6 Gigabike editions are given in Table 2. The results show that the number of players is a parameter that heavily impacts the computation time; at the same time a few hundred players form no computational burden for our model.

¹ More precisely, it would be a multidimensional knapsack problem, given that there is still the constraint on the number of riders in the team.

Table 2 Impact of number of riders on computation time (in seconds)

	100 riders	250 riders	500 riders	750 riders	All riders
Avg. computation time (s)	0.6	6.3	23.8	76.5	200.9

Apart from the standard setting with 6 periods, we also looked at settings with 1, 2 and 3 periods by merging current periods. Notice that a setting with only 1 period does not involve transfers. The average computation times (over 6 Gigabike editions) for various numbers of periods is given in Table 3. Again, we notice that the number of periods has a clear impact on the computation time, though less pronounced than the number of riders. Even without transfers, the problem appears to retain some computational complexity.

Table 3 Impact of number of periods on computation time (in seconds)

	1 period	2 periods	3 periods	6 periods
Avg. computation time (s)	17.6	41.2	44.7	200.9

Despite our experiences with solving the Gigabike instances, which were quite favorable with respect to computation time, we acknowledge that for larger-sized instances of fantasy games (e.g., involving a lot more periods), solving MIP model (1)-(12) optimally could require too much computation time. For such cases a dedicated local search method could be a valuable alternative for the MIP model. Unfortunately, developing an effective local search method that is generally applicable to a broad range of fantasy games is not straightforward. Undoubtedly, the strongest aspect of our MIP model is its generalizability. Ideally, a heuristic method has this property as well. Therefore, we propose a relax-and-fix heuristic. A relax-and-fix heuristic is an iterated search method that enforces the integrality constraints (8)-(11) only for a subset of the variables in MIP model (1)-(12). After each iteration, these variables are fixed to the found optimal values, and a new subset of variables are subject to integrality constraints. The relax-and-fix heuristic proceeds in this way until all variables have an integer value. The main choice that is to be made is how to divide the variables into subsets that will be subsequently fixed. For multi-period fantasy games like Gigabike the different periods represent a natural base for this division. Hence, in the first iteration we only add integrality constraints for all variables of period 1. In the second iteration, the variables of period 1 are fixed to the optimal values found in iteration 1, after which the variables of period 2 are subject to integrality constraints, etc. When applying a relax-and-fix heuristic in this way, we obtained the results reported in the last column of Table 4. From these results we conclude that this relax-and-fix heuristic consistently succeeds in finding near-optimal solutions for the Gigabike game. For the data of 2014, the fix-and-relax heuristic even finds an optimal solution.

Overall, the required computation time for applying the relax-and-fix heuristic on the Gigabike game was only 25% smaller than the time required for solving MIP model (1)-(12). We noticed that the first step in our relax-and-fix heuristic, i.e., a model with integrality constraints on period 1 and no variables fixed, is still relatively hard to solve, as it represents an important part of the whole decision process. Recall that Gigabike consists of six periods and only 1/6 of the riders (i.e., 5 out of 30) can be transferred after each period. Consequently, we expect the efficiency gain to be higher for fantasy games involving a higher number of transfer periods and/or a higher number of transfer opportunities between periods.

4.2 Results

We applied our optimization model to the 2010 through 2015 editions of Gigabike. Our analysis resulted in information value for the organizers as well as the participants, the creation of new subcompetitions and prizes, strategic advice for the participants, and an analysis of the impact of various game rules. The results showed that the Gigabike winner's score accounted for on average 71.6% of the optimal score, leaving a substantial margin for improvement. In fact, the average Gigabike player did not even reach 54% of the optimal score. Optimal scores also allowed us to compare the strength of winners of different editions. Indeed, comparing the scores in absolute terms is not very meaningful, since the total number of points that can be obtained differs from year to year, because of the races that are included (e.g. Olympics in 2012), and

rider's performances (e.g. cheap vs. expensive riders collecting the most points). Computing the optimal score, however, provides a meaningful benchmark to evaluate the participant's performance in Gigabike over the years. Table 4 shows that despite the fact that in absolute terms, the 2012 winner strongly dominates the other winners, the 2010 winner actually performed best and should be called the all-time Gigabike winner.

Table 4: All-time best Gigabike participant

Year	Winner's score	Optimal score	Performance ratio	Relax-and-fix score
2010	26,460	35,064	75.5	34,936
2011	25,347	36,878	68.7	36,836
2012	29,767	40,207	74.0	40,173
2013	26,771	39,122	68.4	39,061
2014	27,374	37,391	73.2	37,391
2015	24,730	35,433	69.8	35,313

We also used our system to compute individualized feedback for each of the participants, in the form of their optimal transfer decisions, given their starting selection. Apart from the fact that many participants found the resulting information highly interesting, and tried to induce strategic insights for the next Gigabike edition, it showed that in fact all participants focusing on the end victory could still (easily) have won the game with the right transfers. Comparing a participant's score with the score that could be obtained given his or her starting team, provided a method to assess the quality of this participant's transfers. Similarly, the quality of the starting selection could be measured. The organizers used this information to give an award to the participants with the best starting team and the best transfers. Upon request, we also computed optimal starting teams and transfers for participants with different goals than the end victory (e.g. winning as many bunch-sprint races as possible).

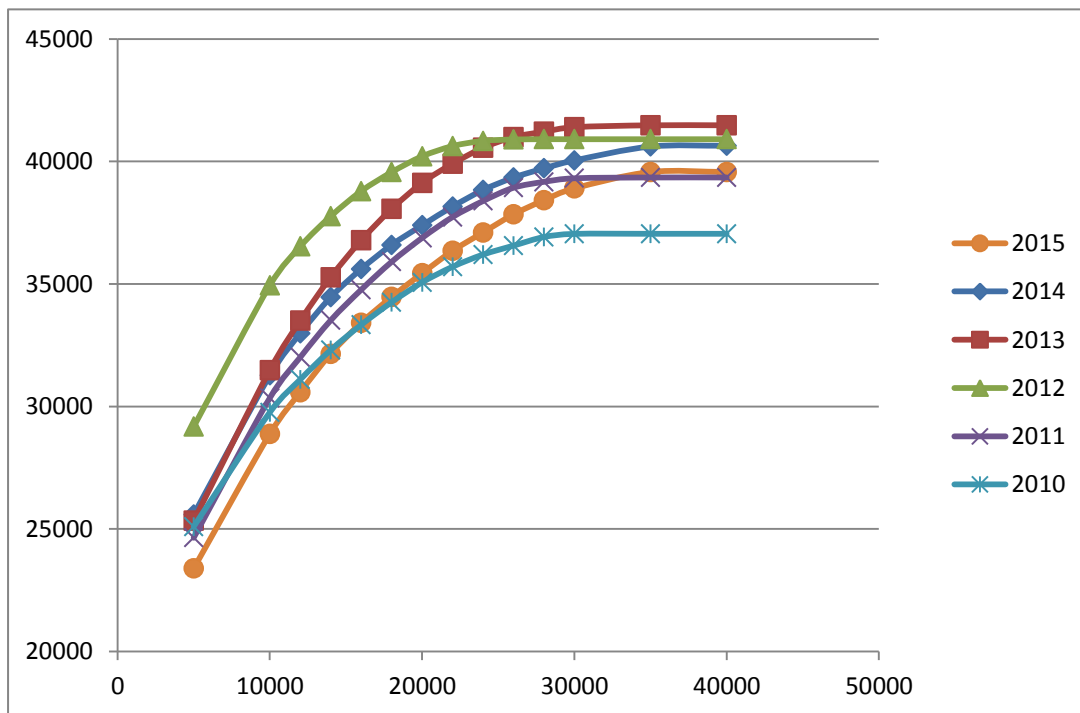


Figure 1 Impact of the budget on the optimal score

A good balance between the importance of the starting team and the transfers had always been a concern for the organizers. More in particular, questions arose as to whether the number of transfer opportunities (5 per period) and the budget (20,000) were adequately chosen. From the individual feedback, we already learned that selecting a starting team was not overly important, in the sense that a poor choice could ruin the rest of the participant's Gigabike experience. However, there is no way of telling what a change of game

rules would trigger with real participants without actually implementing it in at least one edition of Gigabike. Nevertheless, our system allows to compute the effect of rule changes on the optimal score, which we think gives a reasonable idea of the impact it could have on the participants. Figure 1 shows the impact of the budget on the optimal score. The figure shows that the maximal budget best lies in the interval $[10,000 - 30,000]$. Indeed, if the budget is higher than 30,000, the optimal score hardly increases. This means that the budget constraint becomes redundant. With a budget below 10,000 the optimal score dramatically drops. This indicates that the budget restriction strongly forces participants to leave out good performing riders, which turns the game too much into a gambling game in which participants simply have to predict which of the cheap, less known riders will perform better than expected. We conclude that 20,000 is a good choice for the budget restriction. The analysis also reveals that in the interval $[10,000-30,000]$ the value of 1,000 extra budget points is worth around 500 points (on average). Figure 2 depicts the impact of the number of transfer opportunities per period on the optimal score. The figure shows that the value of an extra transfer (slope of the curve) is around 2,000 points if one has only 1 transfer per period. The value decreases with more transfers, however, even with 9 transfers, an extra transfer is still worth around 1,000 points. Based on these results, the choice of 5 transfer opportunities per transfer period seems to be justified.

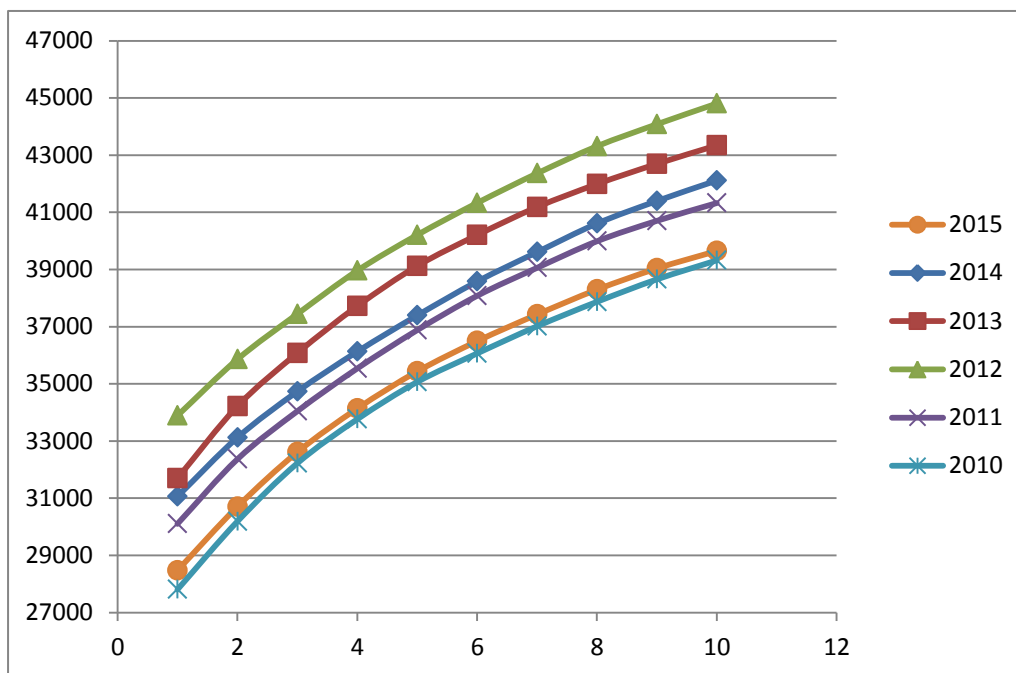


Figure 2 Impact of the number of transfer opportunities on the optimal score.

We performed similar analyses of the number of riders contributing to the team score and the team size, which also confirmed the current settings. Overall, our results managed to provide well-founded support in favor of the current game rules. Moreover, this analysis produced some further strategic advice to the participants. A specific example of such an insight is the trade-off between a number of budget points and an extra transfer opportunity. For instance, participants might face the situation that they have a rider in their team that for sure will lose, say, 1,000 budget points in the next period. This means that this rider could be transferred into the team after the next period for a price of 1,000 budget points below his current price. The decision that the participant has to make is whether or not to transfer this rider out of the team (and back in after the next transfer period) in order to save his budget, at the cost of losing one transfer opportunity. By comparing the curves of Figure 1 and Figure 2, participants have a better idea of the value of an extra transfer compared to the value of extra budget points and, hence, can make a better decision.

A final consideration is whether the budget, which was constant at 20,000 in the last 6 Gigabike editions, has remained equally restrictive over the years. In fact, as the points scored by the riders in the races have (slightly) increased in recent years, the riders have become more expensive. On the other hand, since there

are also more points to collect, this question is hard to answer looking at Figure 1 only. However, the impact of the budget constraint can be measured by the increase in the optimal score when this constraint is removed, as represented in Table 3. The results show that the restrictiveness of the budget is quite similar over the years, given a few exceptions. In 2012, the participants clearly had fewer budget issues to deal with, as many good riders were underpriced because of bad luck or illness in 2011. On the other hand, in 2015, the (expensive) favorites performed exceptionally well, which made budget the prime concern of most participants.

Table 5: Budget restriction over the years

Year	Optimal score	Optimal score (no budget restriction)	Increase (%)
2010	35,064	37,048	8.09
2011	36,878	39,346	6.69
2012	40,207	40,907	1.74
2013	39,122	41,475	6.01
2014	37,391	40,643	8.69
2015	35,433	39,589	11.73

4.3 Impact of uncertainty

To analyze the impact of uncertainty on our findings we performed a scenario analysis on the data of 2012 involving 20 scenarios. In order to obtain realistic scenarios we started from the real results in 2012 and randomly changed the top 10 in each major race (one-day World Tour races and general classifications of each World Tour stage race). As a result, in each scenario the same riders are in the top 10 in each of these races but each time in a different order. Since the points obtained quickly decrease from position 1 to position 10, this random shuffling of top 10's has an important impact on the points collected per rider, but still it reflects a realistic scenario as the difference between winning a race or only finishing in the top 10 is often very small.

In order to evaluate the impact of this uncertainty on our conclusions with respect to the most important game rules, we calculated the optimal scores in each scenario for varying budget limits and varying numbers of transfers per period. In particular, we ran 280 (= 20 scenarios * 14 budget constraints) optimization models for analyzing the impact of the budget and 200 (= 20 scenarios * 10 transfer numbers) optimization models for analyzing the impact of the number of transfers per period.

In order to analyze the results we made a box and whisker plot of the optimal scores found for the 20 scenarios for each setting. These plots can be found in Figure 3 and Figure 4. Figure 3 clearly shows that the maximal budget should be in the interval [10,000 – 30,000]. Of course, the optimal scores vary depending on the scenario, but the box plots clearly show that a budget below 10,000 leads to a dramatic decrease in the optimal score while a budget above 30,000 has (almost) no impact on the optimal score. These results confirm our findings for the deterministic case (see Figure 1).

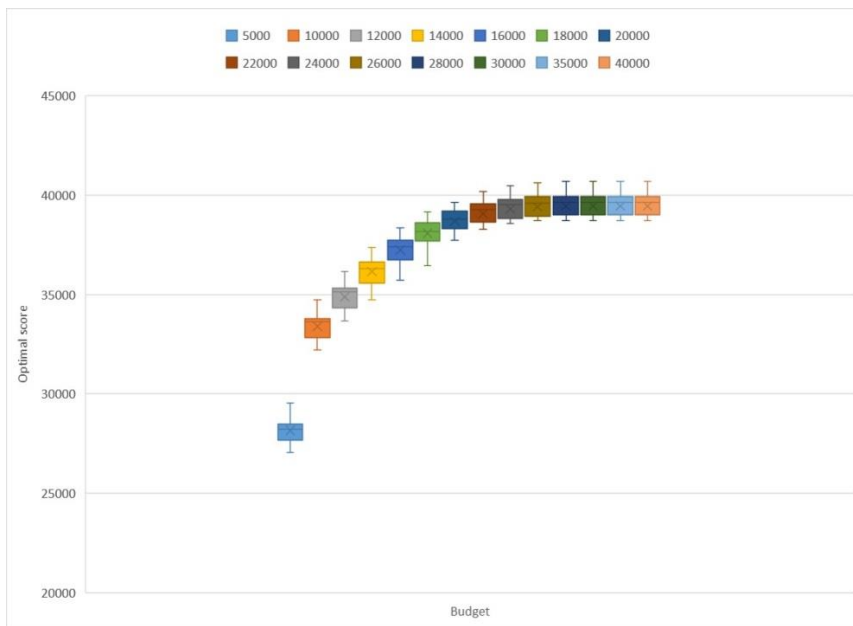


Figure 3 Impact of the budget on the optimal score of Gigabike 2012 when results are uncertain.

Figure 4 clearly shows that also with respect to the number of transfers, our findings based on the deterministic case are confirmed. The value of an extra transfer ranges from 2,000 points if one only has one transfer per period to 1,000 points if one has 9 transfers per period. Another interesting result is the decrease in variability of the optimal score with a growing number of transfers. If the number of transfers is low, the optimal score is highly dependent on the distribution of points over the riders. On the other hand, if the number of transfers is high, the optimal score barely depends on the distribution of points over the riders. A good game tries to avoid both extremes in order to find a good balance between the luck factor, which decreases with a growing number of transfers, and a reasonable range in overall scores of the participants, which increases with a smaller number of transfers. Hence, the stochastic setting confirms our finding that a maximum of 5 transfer opportunities per period is a good game rule.

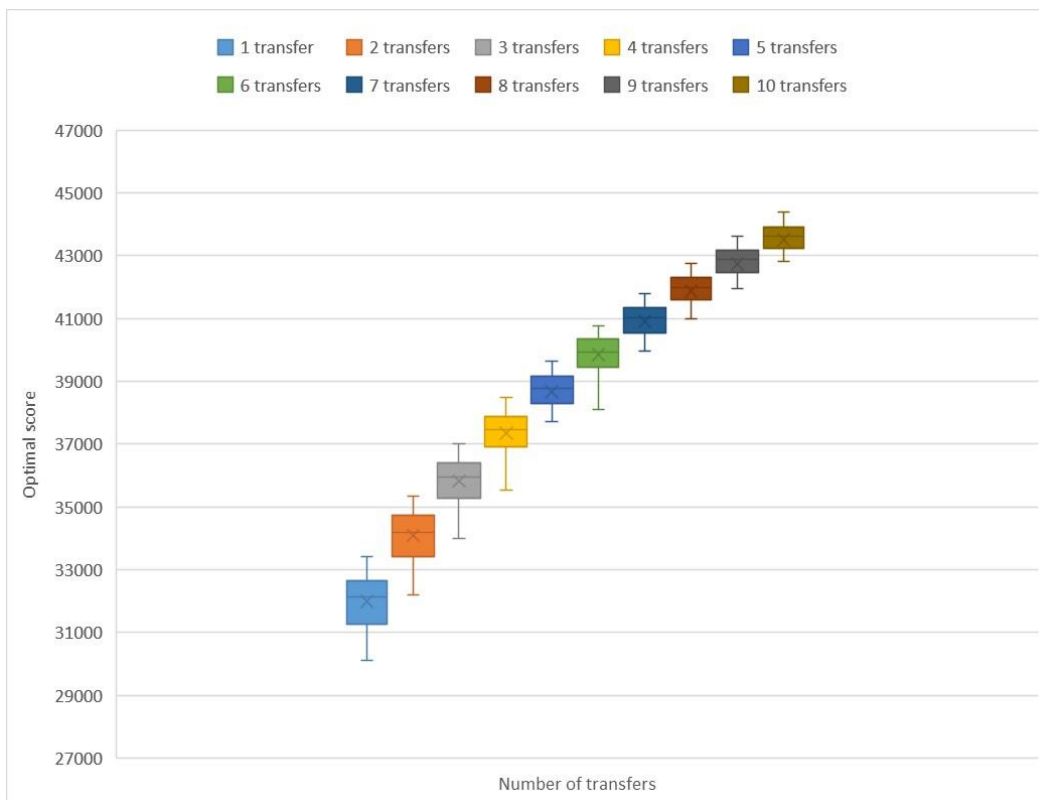


Figure 4 Impact of the number of transfer opportunities on the optimal score of Gigabike 2012 when results are uncertain.

5. Conclusion

We developed an optimization model for fantasy sport games, allowing to compute the best decisions if one knows beforehand how many points each player will collect in each game. We argued that this ex-post optimization model is valuable for both organizers and participants, as it produces informational and commercial value. Based on a survey of fantasy sport games, we listed a number of common characteristics with respect to game rules. Our system is able to deal with each one of them, and hence, is suitable for any fantasy sport game we know.

A practical application to the cycling game Gigabike shows that our system is computationally manageable. Nevertheless, in order to cope with potentially larger-sized instances, we proposed a relax-and-fix heuristic in which variables are fixed per period. This heuristic is not only generally applicable, it also proved to be capable of finding near-optimal results for the Gigabike game. Our application to Gigabike also illustrates how organizers and participants embraced the new possibilities that this model brings. Indeed, the outcome of our model justified the game rules, provided insight in the fantasy game's tactics, allowed for the creation of extra prizes, and increased the joy of participating in the game. A scenario analysis revealed that our findings with respect to game rules and game tactics are robust, i.e. the uncertainty in points collected by each rider leading to different scenarios did not change the conclusions we obtained in the deterministic case.

The fact that our optimization model is valuable for analyzing fantasy sport game rules does not necessarily mean that the model is successful for finding an optimal team and transfers for the ex-ante fantasy sport game, from the perspective of the participants. Although the gained insights from applying the ex-post optimization model, e.g., the value of an extra transfer expressed in budget points, are very useful for the game, a good forecast of the points that will be collected by the players is critical for success. Developing a general stochastic optimization model for fantasy sport games is, however, not straightforward. A stochastic model that maximizes the expected points will most likely not be successful when it comes to obtaining a top rank. The reason is that such a model tends to result in an 'average' team without 'surprising' players, while winning often requires an 'atypical' team with some 'surprising' selections with respect to historical data. A stochastic optimization model that starts from a number of scenarios is also not likely to work well. The number of possible realistic scenarios for a real-life fantasy game, like Gigabike, is so huge, that it is extremely difficult to compose a set of scenarios that can be effectively handled by a stochastic optimization model, like e.g. stochastic programming, and at the same time reflects all possible outcomes reasonably well.

References

- Beliën, J., Goossens, D., Van Reeth, D. & De Boeck, L. (2011). Using mixed integer programming to win a cycling game. *INFORMS Transaction on Education* **11**(3), 93-99.
- Billings, A.C. and Ruihley, B.J. (2013) Why we watch, why we play: the relationship between fantasy sport and fanship motivations. *Mass Communication and Society* **16**, 5-25.
- Bonomo, F., Duran, G., Marengo, J. (2014) Mathematical programming as a tool for virtual soccer coaches: a case study of a fantasy sport game. *International Transactions in Operational Research* **21**, 399-414.
- Dadelo, S., Turskis, Z., Zavadskas, E.K., Dadelienė, R. (2014) Multi-criteria assessment and ranking systems of sport team formation based on objective-measured values of criteria set. *Expert Systems with Applications* **41**, 6106-6113.
- Davis, N.W. and Duncan, M.C. (2006) Sports knowledge is power: reinforcing masculine privilege through fantasy sport league participation. *Journal of Sport and Social Issues* **30**, 244-264.
- Drayer, J., Shapiro, S.L., Dwyer, B., Morse, A.L. and White, J. (2010) The effects of fantasy football participation on NFL consumption: a qualitative analysis. *Sport Management Review* **13**, 129-141.
- Dwyer, B. (2011a) The impact of fantasy football involvement on intentions to watch NFL games on television. *International Journal of Sport Communication* **4**, 375-396.
- Dwyer, B. (2011b) Divided loyalty? An analysis of fantasy football involvement and fan loyalty to individual NFL teams. *Journal of Sport Management* **25**, 445-457.

- Dwyer, B. and Kim, Y. (2011) For love or money: developing and validating a motivational scale for football participation. *Journal of Sport Management* **25**, 70-83.
- Farquhar, L.K. and Meeds, R. (2007) Types of Fantasy Sports Users and Their Motivations. *Journal of computer-mediated communication* **12**, 1208-1228.
- Fortunato, J.A. (2011) The relationship of fantasy football participation with NFL television ratings. *Journal of Sport Administration & Supervision* **3**, 74-90.
- Karg, A.J. and McDonald, H. (2011) Fantasy sport participation as a complement to traditional sport consumption. *Sport Management Review* **14**, 327-346.
- Lee, W.Y., Kwak, D.H., Lim, C., Pedersen, P.M. and Miloch, K.S. (2011) Effects of personality and gender on fantasy sports game participation: the moderating role of perceived knowledge. *Journal of Gambling Studies* **27**, 427-441.
- Lee, S., Seo, W.J. & Green, B.C. (2013) Understanding why people play fantasy sport: development of the Fantasy Sport Motivation Inventory (FanSMI). *European Sport Management Quarterly* **13**, 166-199.
- Levy, D. (2009). Fanship Habitus: The Consumption of Sport in the US. In: Robson, K. & Sanders, C. (eds.) "Quantifying Theory: Pierre Bourdieu", Chapter 15, pp. 187-199, Springer Netherlands.
- Nesbit, T.M. and King, K.A. (2010) The impact of fantasy sports on TV viewership. *Journal of Media Economics* **23**, 24-41.
- Nesbit, T.M. and King, K.A. (2011) Major League Baseball attendance and the role of fantasy baseball. *Journal of Sports Economics* **13**, 494-514.
- Randle, Q. and Nyland, R. (2008) Participation in internet fantasy sports leagues and mass media use. *Journal of website promotion* **3**, 143-152.
- Ruihley, B.J. and Hardin, R.L. (2011) Beyond touchdowns, homeruns, and three-pointers: an examination of fantasy sport motivation. *International Journal of Sport Management & Marketing* **10**, 232-256.
- Sierksma, G. (2006). Computer support for coaching and scouting in football. In: Moritz, E.F. and Haake, S. (eds) *The Engineering of Sport 6, Developments for Innovation* **3**, 215-219. Springer, New York.
- Smith, A.A., Synowka, D.P. and Smith, A.D. (2010) Exploring fantasy sports and its fan base from a CRM perspective. *International Journal of Business Innovation and Research* **4**, 103-142.
- Tavana, M., Azizi, F., Azizi, F. and Behzadian, M. (2013). A fuzzy inference system with application to player selection and team formation in multi-player sports. *Sport Management Review* **16**, 97-110.